
Sequence Listing could not be accepted due to errors.

See attached Validation Report.

If you need help call the Patent Electronic Business Center at (866)

217-9197 (toll free).

Reviewer: Durreshwar Anjum

Timestamp: [year=2008; month=5; day=30; hr=12; min=17; sec=48; ms=565;]

Reviewer Comments:

<210> 30

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 30

gcttatggcg cgcctcagga ctggagcttg ctccgc

36n

21

Please delete the extra n which appears beside the total number of bases 36 and also remove the end of file text which is below the bases (21) as shown in the attachment above.

Validated By CRFValidator v 1.0.3

Application No: 10590956 Version No: 1.0

Input Set:

Output Set:

Started: 2008-05-09 16:08:23.466

Finished: 2008-05-09 16:08:25.990

Elapsed: 0 hr(s) 0 min(s) 2 sec(s) 524 ms

Total Warnings: 24

Total Errors: 7

No. of SeqIDs Defined: 30

Actual SeqID Count: 30

Error code	Error Description
W 402	Undefined organism found in <213> in SEQ ID (1)
W 402	Undefined organism found in <213> in SEQ ID (2)
W 402	Undefined organism found in <213> in SEQ ID (3)
W 402	Undefined organism found in <213> in SEQ ID (4)
W 402	Undefined organism found in <213> in SEQ ID (5)
W 402	Undefined organism found in <213> in SEQ ID (6)
W 213	Artificial or Unknown found in <213> in SEQ ID (13)
W 213	Artificial or Unknown found in <213> in SEQ ID (14)
W 213	Artificial or Unknown found in <213> in SEQ ID (15)
W 213	Artificial or Unknown found in <213> in SEQ ID (16)
W 213	Artificial or Unknown found in <213> in SEQ ID (17)
W 213	Artificial or Unknown found in <213> in SEQ ID (18)
W 213	Artificial or Unknown found in <213> in SEQ ID (19)
W 402	Undefined organism found in <213> in SEQ ID (20)
W 213	Artificial or Unknown found in <213> in SEQ ID (21)
W 213	Artificial or Unknown found in <213> in SEQ ID (22)
W 213	Artificial or Unknown found in <213> in SEQ ID (23)
W 213	Artificial or Unknown found in <213> in SEQ ID (24)
W 213	Artificial or Unknown found in <213> in SEQ ID (25)
W 213	Artificial or Unknown found in <213> in SEQ ID (26)

Input Set:

Output Set:

Started: 2008-05-09 16:08:23.466

Finished: 2008-05-09 16:08:25.990

Elapsed: 0 hr(s) 0 min(s) 2 sec(s) 524 ms

Total Warnings: 24

Total Errors: 7

No. of SeqIDs Defined: 30

Actual SeqID Count: 30

Error code		Error Description
W	213	Artificial or Unknown found in <213> in SEQ ID (27)
M	213	Artificial or Unknown found in <213> in SEQ ID (28)
W	213	Artificial or Unknown found in <213> in SEQ ID (29)
W	213	Artificial or Unknown found in <213> in SEQ ID (30)
E	342	'n' position not defined found at POS: 39 SEQID(30)
E	254	The total number of bases conflicts with running total Input: 0, Calculated: 39 SEQID(30)
E	323	Invalid/missing amino acid numbering SEQID (30)at Protein (1)
E	323	Invalid/missing amino acid numbering SEQID (30) POS (1)
E	323	Invalid/missing amino acid numbering SEQID (30)at Protein (5)
E	323	Invalid/missing amino acid numbering SEQID (30)at Protein (10)
E	253	The number of bases differs from <211> Input: 36 Calculated:39

SEQUENCE LISTING

```
<110> Genencor International, Inc.
      Bower, Benjamin
      Mitchinson, Colin
      Larenas, Edmund
<120> Cellulase Fusion Protein and Heterologous Cellulase Fusion
      Construct Encoding the Same
<130> GC832-PCT
<140> 10590956
<141> 2008-05-09
<150> PCT/US2005/010242
<151> 2005-03-25
<150> US 60/556,711
<151> 2004-03-25
<160> 30
<170> PatentIn version 3.2
<210> 1
<211> 1570
<212> DNA
<213> Trichoderma reesei
<400> 1
atgtategga agttggeegt cateteggee ttettggeea eagetegtge teagteggee
                                                                     60
tgcactetee aateggagae teaecegeet etgacatgge agaaatgete gtetggtgge
                                                                     120
acttgcactc aacagacagg ctccgtggtc atcgacgcca actggcgctg gactcacgct
                                                                    180
acquacagca gcacquactg ctacquatggc accacttgga gctcgaccct atgtcctgac
                                                                    240
aacgagacct gcgcgaagaa ctgctgtctg gacggtgccg cctacgcgtc cacgtacgga
                                                                    300
gttaccacga gcggtaacag cctctccatt ggctttgtca cccagtctgc gcagaagaac
                                                                     360
                                                                     420
gttggcgctc gcctttacct tatggcgagc gacacgacct accaggaatt caccctgctt
ggcaacgagt tctctttcga tgttgatgtt tcgcagctgc cgtaagtgac ttaccatgaa
                                                                     480
cccctgacgt atcttcttgt gggctcccag ctgactggcc aatttaaggt gcggcttgaa
                                                                     540
cggagctctc tacttcgtgt ccatggacgc ggatggtggc gtgagcaagt atcccaccaa
                                                                     600
caccgctggc gccaagtacg gcacggggta ctgtgacagc cagtgtcccc gcgatctgaa
                                                                     660
                                                                    720
gttcatcaat ggccaggcca acgttgaggg ctgggagccg tcatccaaca acgcaaacac
                                                                    780
gggcattgga ggacacggaa gctgctgctc tgagatggat atctgggagg ccaactccat
ctccgaggct cttacccccc acccttgcac gactgtcggc caggagatct gcgagggtga
                                                                    840
                                                                    900
tgggtgcggc ggaacttact ccgataacag atatggcggc acttgcgatc ccgatggctg
cgactggaac ccataccgcc tgggcaacac cagcttctac ggccctggct caagctttac
                                                                    960
cctcgatacc accaagaaat tgaccgttgt cacccagttc gagacgtcgg gtgccatcaa
                                                                  1020
ccgatactat gtccagaatg gcgtcacttt ccagcagccc aacgccgagc ttggtagtta
                                                                   1080
ctctggcaac gagctcaacg atgattactg cacagctgag gaggcagaat tcggcggatc
                                                                   1140
ctctttctca gacaagggcg gcctgactca gttcaagaag gctacctctg gcggcatggt
                                                                   1200
tctggtcatg agtctgtggg atgatgtgag tttgatggac aaacatgcgc gttgacaaag
                                                                    1260
agtcaagcag ctgactgaga tgttacagta ctacgccaac atgctgtggc tggactccac
                                                                   1320
                                                                    1380
ctacccgaca aacgagacct cctccacacc cggtgccgtg cgcggaagct gctccaccag
ctccggtgtc cctgctcagg tcgaatctca gtctcccaac gccaaggtca ccttctccaa
                                                                    1440
catcaagttc ggacccattg gcagcaccgg caaccctagc ggcggcaacc ctcccggcgg
                                                                    1500
```

aaacccgcct ggcaccacca ccacccgccg cccagccact accactggaa gctctcccgg

1560

```
acctactagt
                                                                    1570
<210>
<211>
      51
<212>
      DNA
<213> Trichoderma reesei
<400> 2
atgtatcgga agttggccgt catctcggcc ttcttggcca cagctcgtgc t
                                                                      51
<210> 3
<211> 1438
<212> DNA
<213>
      Trichoderma reesei
<400>
                                                                      60
cagteggeet geacteteea ateggagaet caeeegeete tgacatggea gaaatgeteg
tetggtggca ettgcaetca acagacagge teegtggtea tegaegeeaa etggegetgg
                                                                     120
actcacgcta cgaacagcag cacgaactgc tacgatggca acacttggag ctcgacccta
                                                                     180
                                                                     240
tgtcctgaca acgagacctg cgcgaagaac tgctgtctgg acggtgccgc ctacgcgtcc
acgtacggag ttaccacgag cggtaacagc ctctccattg gctttgtcac ccagtctgcg
                                                                     300
                                                                     360
cagaagaacg ttggcgctcg cctttacctt atggcgagcg acacgaccta ccaggaattc
accetgettg geaacgagtt etetttegat gttgatgttt egeagetgee gtaagtgaet
                                                                     420
taccatgaac ccctgacgta tcttcttgtg ggctcccagc tgactggcca atttaaggtg
                                                                     480
                                                                     540
cggcttgaac ggagctctct acttcgtgtc catggacgcg gatggtggcg tgagcaagta
teccaecaae acegetggeg ecaagtaegg eaeggggtae tgtgaeagee agtgteeeeg
                                                                     600
                                                                     660
cgatctgaag ttcatcaatg gccaggccaa cgttgagggc tgggagccgt catccaacaa
cgcaaacacg ggcattggag gacacggaag ctgctgctct gagatggata tctgggaggc
                                                                     720
                                                                     780
caactccatc teegaggete ttaceeecca ecettgeacg actgteggee aggagatetg
                                                                     840
cgagggtgat gggtgcggcg gaacttactc cgataacaga tatggcggca cttgcgatcc
cgatggctgc gactggaacc cataccgcct gggcaacacc agcttctacg gccctggctc
                                                                     900
                                                                     960
aagctttacc ctcgatacca ccaagaaatt gaccgttgtc acccagttcg agacgtcggg
tgccatcaac cgatactatg tccagaatgg cgtcactttc cagcagccca acgccgagct
                                                                    1020
                                                                    1080
tggtagttac tctggcaacg agctcaacga tgattactgc acagctgagg aggcagaatt
cggcggatcc tctttctcag acaagggcgg cctgactcag ttcaagaagg ctacctctgg
                                                                    1140
cggcatggtt ctggtcatga gtctgtggga tgatgtgagt ttgatggaca aacatgcgcg
                                                                    1200
ttgacaaaga gtcaagcagc tgactgagat gttacagtac tacgccaaca tgctgtggct
                                                                    1260
ggactccacc tacccgacaa acgagacctc ctccacaccc ggtgccgtgc gcggaagctg
                                                                    1320
                                                                    1380
ctccaccage teeggtgtee etgeteaggt egaateteag teteceaaeg ecaaggteae
cttctccaac atcaagttcg gacccattgg cagcaccggc aaccctagcg gcggcaac
                                                                    1438
<210>
      4
<211> 81
<212> DNA
<213> Trichoderma reesei
<400>
ceteceggeg gaaaceegee tggcaceace accaeeegee geeeageeac taccaetgga
                                                                      60
agctctcccg gacctactag t
                                                                      81
<210>
      5
<211>
      480
<212>
      PRT
<213>
      Trichoderma reesei
```

<400>

5

Met Tyr 1	Arg	Lys	Leu 5	Ala	Val	Ile	Ser	Ala 10	Phe	Leu	Ala	Thr	Ala 15	Arg
Ala Gln	Ser	Ala 20	Cys	Thr	Leu	Gln	Ser 25	Glu	Thr	His	Pro	Pro 30	Leu	Thr
Trp Gln	Lys 35	Cys	Ser	Ser	Gly	Gly 40	Thr	Cys	Thr	Gln	Gln 45	Thr	Gly	Ser
Val Val 50	Ile	Asp	Ala	Asn	Trp 55	Arg	Trp	Thr	His	Ala 60	Thr	Asn	Ser	Ser
Thr Asn 65	Cys	Tyr	Asp	Gly 70	Asn	Thr	Trp	Ser	Ser 75	Thr	Leu	Cys	Pro	Asp 80
Asn Glu	Thr	Cys	Ala 85	Lys	Asn	Cys	Cys	Leu 90	Asp	Gly	Ala	Ala	Tyr 95	Ala
Ser Thr	Tyr	Gly 100	Val	Thr	Thr	Ser	Gly 105	Asn	Ser	Leu	Ser	Ile 110	Gly	Phe
Val Thr	Gln 115	Ser	Ala	Gln	Lys	Asn 120	Val	Gly	Ala	Arg	Leu 125	Tyr	Leu	Met
Ala Ser 130	Asp	Thr	Thr	Tyr	Gln 135	Glu	Phe	Thr	Leu	Leu 140	Gly	Asn	Glu	Phe
Ser Phe 145	Asp	Val	Asp	Val 150	Ser	Gln	Leu	Pro	Cys 155	Gly	Leu	Asn	Gly	Ala 160
Leu Tyr	Phe	Val	Ser 165	Met	Asp	Ala	Asp	Gly 170	Gly	Val	Ser	Lys	Tyr 175	Pro
Thr Asn	Thr	Ala 180	Gly	Ala	Lys	Tyr	Gly 185	Thr	Gly	Tyr	Cys	Asp 190	Ser	Gln
Cys Pro	Arg 195	Asp	Leu	Lys	Phe	Ile 200	Asn	Gly	Gln	Ala	Asn 205	Val	Glu	Gly
Trp Glu 210	Pro	Ser	Ser	Asn	Asn 215	Ala	Asn	Thr	Gly	Ile 220	Gly	Gly	His	Gly
Ser Cys	Cys	Ser	Glu		Asp	Ile	Trp	Glu		Asn	Ser	Ile	Ser	
225	m1	_		230	~	m1	 1	7	235	~ 1	~ 1	1	~	240
Ala Leu			245		_			250	_				255	
Gly Asp	Gly	Cys 260	Gly	Gly	Thr	Tyr	Ser 265	Asp	Asn	Arg	Tyr	Gly 270	Gly	Thr
Cys Asp		Asp	Gly	Cys	Asp	_	Asn	Pro	Tyr	Arg	Leu 285	Gly	Asn	Thr
	275					280					200			
Ser Phe 290	Tyr	_		_	295	Ser				300	Thr		_	_
290 Leu Thr	Tyr	_		Gln	295	Ser			Gly	300	Thr		_	Tyr
290 Leu Thr 305	Tyr Val	Val	Thr	Gln 310	295 Phe	Ser Glu	Thr	Ser	Gly 315	300 Ala	Thr	Asn	Arg	Tyr 320
290 Leu Thr	Tyr Val	Val	Thr Gly	Gln 310	295 Phe	Ser Glu	Thr	Ser Gln	Gly 315	300 Ala	Thr	Asn	Arg Leu	Tyr 320
290 Leu Thr 305 Tyr Val	Tyr Val Gln	Val Asn	Thr Gly 325	Gln 310 Val	295 Phe Thr	Ser Glu Phe	Thr Gln	Ser Gln 330	Gly 315 Pro	300 Ala Asn	Thr Ile Ala	Asn Glu	Arg Leu 335	Tyr 320 Gly
290 Leu Thr 305	Tyr Val Gln	Val Asn	Thr Gly 325	Gln 310 Val	295 Phe Thr	Ser Glu Phe	Thr Gln	Ser Gln 330	Gly 315 Pro	300 Ala Asn	Thr Ile Ala	Asn Glu	Arg Leu 335	Tyr 320 Gly
290 Leu Thr 305 Tyr Val	Tyr Val Gln Ser	Val Asn Gly 340	Thr Gly 325 Asn	Gln 310 Val Glu	295 Phe Thr Leu	Ser Glu Phe Asn	Thr Gln Asp 345	Ser Gln 330 Asp	Gly 315 Pro	300 Ala Asn Cys	Thr Ile Ala Thr	Asn Glu Ala 350	Arg Leu 335 Glu	Tyr 320 Gly
290 Leu Thr 305 Tyr Val Ser Tyr	Tyr Val Gln Ser Phe 355	Val Asn Gly 340 Gly	Thr Gly 325 Asn Gly	Gln 310 Val Glu Ser	295 Phe Thr Leu Ser	Ser Glu Phe Asn Phe 360	Thr Gln Asp 345 Ser	Ser Gln 330 Asp	Gly 315 Pro Tyr	300 Ala Asn Cys	Thr Ile Ala Thr Gly 365	Asn Glu Ala 350 Leu	Arg Leu 335 Glu Thr	Tyr 320 Gly Glu
Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys	Tyr Val Gln Ser Phe 355 Lys	Val Asn Gly 340 Gly Ala	Thr Gly 325 Asn Gly Thr	Gln 310 Val Glu Ser	295 Phe Thr Leu Ser Gly 375	Ser Glu Phe Asn Phe 360 Gly	Thr Gln Asp 345 Ser Met	Ser Gln 330 Asp Asp	Gly 315 Pro Tyr Lys	300 Ala Asn Cys Gly Val 380	Thr Ile Ala Thr Gly 365 Met	Asn Glu Ala 350 Leu Ser	Arg Leu 335 Glu Thr	Tyr 320 Gly Glu Gln
290 Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys 370 Asp Asp	Tyr Val Gln Ser Phe 355 Lys Tyr	Val Asn Gly 340 Gly Ala	Thr Gly 325 Asn Gly Thr	Gln 310 Val Glu Ser Ser Asn 390	295 Phe Thr Leu Ser Gly 375 Met	Ser Glu Phe Asn Phe 360 Gly Leu	Thr Gln Asp 345 Ser Met Trp	Ser Gln 330 Asp Asp Val Leu	Gly 315 Pro Tyr Lys Leu Asp 395	300 Ala Asn Cys Gly Val 380 Ser	Thr Ile Ala Thr Gly 365 Met Thr	Asn Glu Ala 350 Leu Ser	Arg Leu 335 Glu Thr Leu Pro	Tyr 320 Gly Glu Gln Trp Thr 400
Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys 370 Asp Asp 385	Tyr Val Gln Ser Phe 355 Lys Tyr Thr	Val Asn Gly 340 Gly Ala Tyr Ser Val	Thr Gly 325 Asn Gly Thr Ala Ser 405	Gln 310 Val Glu Ser Ser Asn 390 Thr	295 Phe Thr Leu Ser Gly 375 Met	Ser Glu Phe Asn Phe 360 Gly Leu Gly	Thr Gln Asp 345 Ser Met Trp Ala Glu	Ser Gln 330 Asp Val Leu Val 410	Gly 315 Pro Tyr Lys Leu Asp 395 Arg	300 Ala Asn Cys Gly Val 380 ser	Thr Ile Ala Thr Gly 365 Met Thr	Asn Glu Ala 350 Leu Ser Tyr Cys Asn	Arg Leu 335 Glu Thr Leu Pro Ser 415	Tyr 320 Gly Glu Gln Trp Thr 400 Thr
290 Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys 370 Asp Asp 385 Asn Glu Ser Ser	Tyr Val Gln Ser Phe 355 Lys Tyr Thr Gly	Val Asn Gly 340 Gly Ala Tyr Ser Val 420	Thr Gly 325 Asn Gly Thr Ala Ser 405 Pro	Gln 310 Val Glu Ser Ser Asn 390 Thr	295 Phe Thr Leu Ser Gly 375 Met Pro Gln	Ser Glu Phe Asn Phe 360 Gly Leu Gly Val	Thr Gln Asp 345 Ser Met Trp Ala Glu 425	Ser Gln 330 Asp Val Leu Val 410 Ser	Gly 315 Pro Tyr Lys Leu Asp 395 Arg	300 Ala Asn Cys Gly Val 380 ser Gly Ser	Thr Ile Ala Thr Gly 365 Met Thr Ser	Asn Glu Ala 350 Leu Ser Tyr Cys Asn 430	Arg Leu 335 Glu Thr Leu Pro Ser 415 Ala	Tyr 320 Gly Glu Gln Trp Thr 400 Thr
290 Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys 370 Asp Asp 385 Asn Glu	Tyr Val Gln Ser Phe 355 Lys Tyr Thr Gly Phe	Val Asn Gly 340 Gly Ala Tyr Ser Val 420	Thr Gly 325 Asn Gly Thr Ala Ser 405 Pro	Gln 310 Val Glu Ser Ser Asn 390 Thr	295 Phe Thr Leu Ser Gly 375 Met Pro Gln	Ser Glu Phe Asn Phe 360 Gly Leu Gly Val Phe	Thr Gln Asp 345 Ser Met Trp Ala Glu 425	Ser Gln 330 Asp Val Leu Val 410 Ser	Gly 315 Pro Tyr Lys Leu Asp 395 Arg	300 Ala Asn Cys Gly Val 380 ser Gly Ser	Thr Ile Ala Thr Gly 365 Met Thr Ser Pro	Asn Glu Ala 350 Leu Ser Tyr Cys Asn 430	Arg Leu 335 Glu Thr Leu Pro Ser 415 Ala	Tyr 320 Gly Glu Gln Trp Thr 400 Thr
290 Leu Thr 305 Tyr Val Ser Tyr Ala Glu Phe Lys 370 Asp Asp 385 Asn Glu Ser Ser	Tyr Val Gln Ser Phe 355 Lys Tyr Thr Gly Phe 435	Val Asn Gly 340 Gly Ala Tyr Ser Val 420 Ser	Thr Gly 325 Asn Gly Thr Ala Ser 405 Pro	Gln 310 Val Glu Ser Ser Asn 390 Thr Ala	295 Phe Thr Leu Ser Gly 375 Met Pro Gln Lys	Ser Glu Phe Asn Phe 360 Gly Leu Gly Val Phe 440	Thr Gln Asp 345 Ser Met Trp Ala Glu 425 Gly	Ser Gln 330 Asp Val Leu Val 410 Ser	Gly 315 Pro Tyr Lys Leu Asp 395 Arg Gln Ile	300 Ala Asn Cys Gly Val 380 Ser Gly Ser	Thr Ile Ala Thr Gly 365 Met Thr Ser Pro Ser 445	Asn Glu Ala 350 Leu Ser Tyr Cys Asn 430 Thr	Arg Leu 335 Glu Thr Leu Pro Ser 415 Ala Gly	Tyr 320 Gly Glu Gln Trp Thr 400 Thr Lys

450 455 460

Thr Arg Arg Pro Ala Thr Thr Gly Ser Ser Pro Gly Pro Thr Ser 465 470 475 480

<210> 6

<211> 431

<212> PRT

<213> Trichoderma reesei

<400> 6

Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr Trp

1 5 10 15

Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser Val 20 25 30

Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser Thr
35 40 45

Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp Asn 50 55 60

Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala Ser 65 70 75 80

Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe Val
85 90 95

Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met Ala 100 105 110

Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe Ser 115 120 125

Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala Leu 130 135 140

Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Thr 145 150 155 160

Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys 165 170 175

Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly Trp 180 185 190

Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly Ser 195 200 205

Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu Ala 210 215 220

Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu Gly
225 230 235 240

Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr Cys
245 250 255

Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr Ser 260 265 270

Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys Leu 275 280 285

Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr Tyr

290 295 300 Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly Ser

305 310 315 320 Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu Ala 325 330 335

Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln Phe 340 345 350

Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp Asp 355 360 365

Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr Asn

370 375 380 Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr Ser 385 390 395 400 Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys Val 405 410 Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn 420 425 <210> 7 <211> 1077 <212> DNA <213> Acidothermus cellulolyticus <400> qcqqqcqqcq qctattqqca cacqaqcqqc cqqqaqatcc tqqacqcqaa caacqtqccq 60 gtacggatcg ccggcatcaa ctggtttggg ttcgaaacct gcaattacgt cgtgcacggt 120 ctctggtcac gcgactaccg cagcatgctc gaccagataa agtcgctcgg ctacaacaca 180 atccggctgc cgtactctga cgacattctc aagccgggca ccatgccgaa cagcatcaat 240 ttttaccaga tgaatcagga cctgcagggt ctgacgtcct tgcaggtcat ggacaaaatc 300 gtcgcgtacg ccggtcagat cggcctgcgc atcattcttg accgccaccg accggattgc 360 agegggeagt eggegetgtg gtacaegage agegtetegg aggetaegtg gattteegae 420 480 ctgcaagcgc tggcgcagcg ctacaaggga aacccgacgg tcgtcggctt tgacttgcac 540 aacgagccgc atgacccggc ctgctggggc tgcggcgatc cgagcatcga ctggcgattg 600 gccgccgagc gggccggaaa cgccgtgctc tcggtgaatc cgaacctgct cattttcgtc gaaggtgtgc agagctacaa cggagactcc tactggtggg gcggcaacct gcaaggagcc 660 ggccagtacc cggtcgtgct gaacgtgccg aaccgcctgg tgtactcggc gcacgactac 720 gcgacgagcg tctacccgca gacgtggttc agcgatccga ccttccccaa caacatgccc 780 840 ggcatctgga acaagaactg gggatacctc ttcaatcaga acattgcacc ggtatggctg ggcgaattcg gtacgacact gcaatccacg accgaccaga cgtggctgaa gacgctcgtc 900 cagtacctac ggccgaccgc gcaatacggt gcggacagct tccagtggac cttctggtcc 960 tggaaccccg attccggcga cacaggagga attctcaagg atgactggca gacggtcgac 1020 acagtaaaag acggctatct cgcgccgatc aagtcgtcga ttttcgatcc tgtcggc 1077 <210> 8 <211> 359 <212> PRT <213> Acidothermus cellulolyticus <400> 8 Ala Gly Gly Gly Tyr Trp His Thr Ser Gly Arg Glu Ile Leu Asp Ala Asn Asn Val Pro Val Arg Ile Ala Gly Ile Asn Trp Phe Gly Phe Glu 20 2.5 30 Thr Cys Asn Tyr Val Val His Gly Leu Trp Ser Arg Asp Tyr Arg Ser 40 45 Met Leu Asp Gln Ile Lys Ser Leu Gly Tyr Asn Thr Ile Arg Leu Pro 55 Tyr Ser Asp Asp Ile Leu Lys Pro Gly Thr Met Pro Asn Ser Ile Asn 70 75 Phe Tyr Gln Met Asn Gln Asp Leu Gln Gly Leu Thr Ser Leu Gln Val 90 Met Asp Lys Ile Val Ala Tyr Ala Gly Gln Ile Gly Leu Arg Ile Ile 100 105 Leu Asp Arg His Arg Pro Asp Cys Ser Gly Gln Ser Ala Leu Trp Tyr 120 125

Thr Ser Ser Val Ser Glu Ala Thr Trp Ile Ser Asp Leu Gln Ala Leu

140

135

130

```
Ala Gln Arg Tyr Lys Gly Asn Pro Thr Val Val Gly Phe Asp Leu His
   150
                              155 160
Asn Glu Pro His Asp Pro Ala Cys Trp Gly Cys Gly Asp Pro Ser Ile
      165
                           170
Asp Trp Arg Leu Ala Ala Glu Arg Ala Gly Asn Ala Val Leu Ser Val
               185
        180
Asn Pro Asn Leu Leu Ile Phe Val Glu Gly Val Gln Ser Tyr Asn Gly
      195 200
Asp Ser Tyr Trp Trp Gly Gly Asn Leu Gln Gly Ala Gly Gln Tyr Pro
                  215
Val Val Leu Asn Val Pro Asn Arg Leu Val Tyr Ser Ala His Asp Tyr
225 230 235 240
Ala Thr Ser Val Tyr Pro Gln Thr Trp Phe Ser Asp Pro Thr Phe Pro
                   250
Asn Asn Met Pro Gly Ile Trp Asn Lys Asn Trp Gly Tyr Leu Phe Asn
         260 265
Gln Asn Ile Ala Pro Val Trp Leu Gly Glu Phe Gly Thr Thr Leu Gln
      275 280
Ser Thr Thr Asp Gln Thr Trp Leu Lys Thr Leu Val Gln Tyr Leu Arg
                   295
                                   300
Pro Thr Ala Gln Tyr Gly Ala Asp Ser Phe Gln Trp Thr Phe Trp Ser
                               315
               310
Trp Asn Pro Asp Ser Gly Asp Thr Gly Gly Ile Leu Lys Asp Asp Trp
                   330
           325
Gln Thr Val Asp Thr Val Lys Asp Gly Tyr Leu Ala Pro Ile Lys Ser
                        345
Ser Ile Phe Asp Pro Val Gly
     355
<210> 9
<211> 1914
<212> DNA
<213> Acidothermus cellulolyticus
```

<400> 9

aacgacccgt	acatccagcg	gttcctcacg	atgtacaaca	agattcacga	cccagcgaac	60
ggctacttca	gcccgcaggg	aattccctac	cactcggtag	aaacgctcat	cgttgaggca	120
ccggactacg	ggcacgagac	aacttcggag	gcgtacagct	tctggctctg	gctcgaagcg	180
acgtacggcg	cagtgaccgg	caactggacg	ccgttcaaca	acgcctggac	gacgatggaa	240
acgtacatga	tcccgcagca	cgcggaccag	ccgaacaacg	cgtcgtacaa	ccccaacagc	300
ccggcgtcgt	acgctccgga	agagccgctg	cccagcatgt	acccggttgc	catcgacagc	360
agcgtgccgg	ttgggcacga	cccgctcgcc	gccgaattgc	agtcgacgta	cggcactccg	420
gacatttacg	gcatgcactg	gctggccgac	gttgacaaca	tctacggata	cggcgacagc	480
cccggcggtg	gttgcgaact	cggtccttcc	gctaagggcg	tctcctacat	caacacattc	540
cagcgcggct	cgcaggagtc	cgtctgggag	acggtcaccc	agccgacgtg	cgacaacggc	600
aagtacggtg	gggcgcacgg	ctacgtcgac	ctgttcatcc	agggttcgac	gccgccgcag	660
tggaagtaca	ccgatgcccc	ggacgccgac	gcccgtgccg	tccaggctgc	gtactgggcc	720
tacacctg						